Examining the Distribution of Experimental Averages of Samples of the Exponential Distribution

Sean Dunagan

September 25, 2015

For this project, we generated samples consisting of 40 random number generations (hereby referred to as observations) distributed via the exponential distribution. We then took the mean value among the 40 observations in each sample. It is these 40-observations-means (hereby referred to as averages) that we examined, calculating the mean and variance of the distribution of these averages. We then compared the distribution of these averages to a normal distribution whose mean and variance matched the theoretical mean and variance of our set of averages.

Function generateBatchesOfObservations() was the function which actually executed the random number generation. Function rexp() was used to generate the observations. The mean() function was used to evaluate the averages of each set of 40 observations. mean() was also used to generate the expiremental mean of the averages; sd() was used to compute the standard deviation of the averages (and therefore the variance as well). Function produceStatisticsReport() output the theoretical and experimental values associated with the generated observation averages. Function produceSampleDensityHistogramAgainstNormalCurve() created a histogram overlaying the theoretical normal distribution of observation averages over our set of experimental observation averages. In the histogram, our experimental oversation averages are displayed as densities, not as frequencies; this allows us to easily use dnorm() to generate the normal curve necessary for comparison. The code is commented to explain step-by-step how these plots are generated.

The Theoretical mean of the Sample Averages is the population mean, which is 1/lambda for the exponential distribution. As such, with lambda fixed to .2 for our simulations, the population mean was 5, resulting in the theoretical sample average to also have an expected mean of 5.

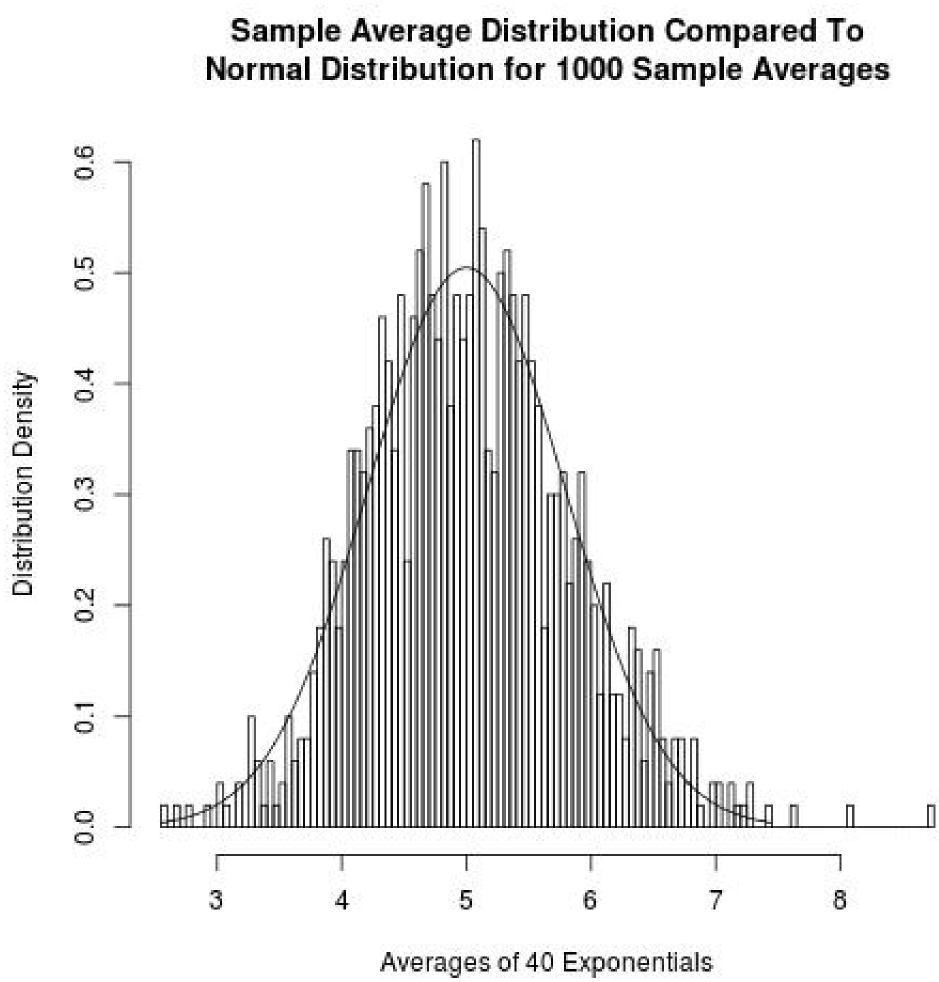
The Theoretical Population Variance refers to the amount of variation we expect in a random variable. This value is independent of the number of observations which may be sampled from a population. The Theoretical Variance of the Sample Averages is the Population Variance divided by the number of observations n which comprise each Average’s sample size. As such, the more observations we have comprising each sample average, the less we expect the distribution of the sample averages to vary; this is due to the law of large numbers taking effect. This differs from the population variance because it refers to the distribution of experimental sample AVERAGES, not the distribution of individual members of the population. Examining the distribution of averages allows us to smooth away outliers in the data population.

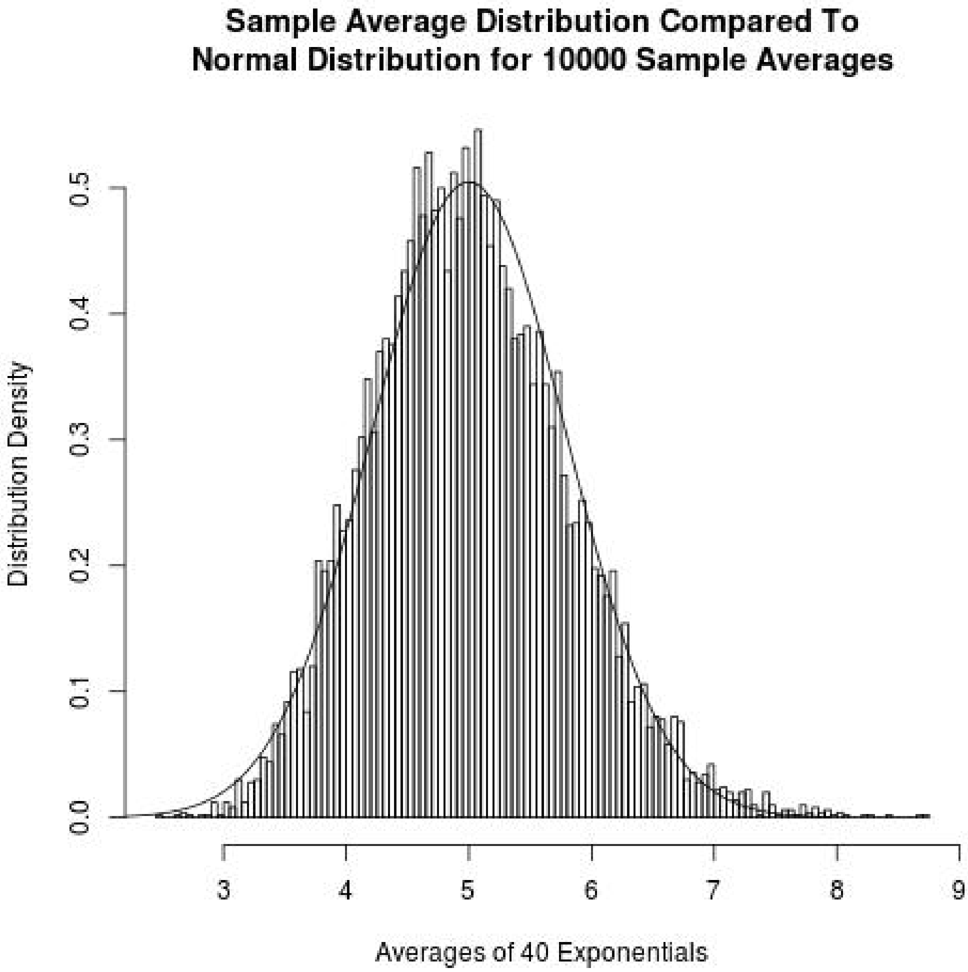
As you can see below, while the experimental mean approximated the population mean for all sample sizes, the variance decreased as the size of the samples increased. Theoretically this is expected as the theoretical variance of a sample mean is the population mean divided by the number of observations per sample, mentioned above. Intuitively this makes sense as the law of large numbers implies that as we take more samples, we expect that the average observation should approach the expected value. This is why the expected value is expected in the first place.

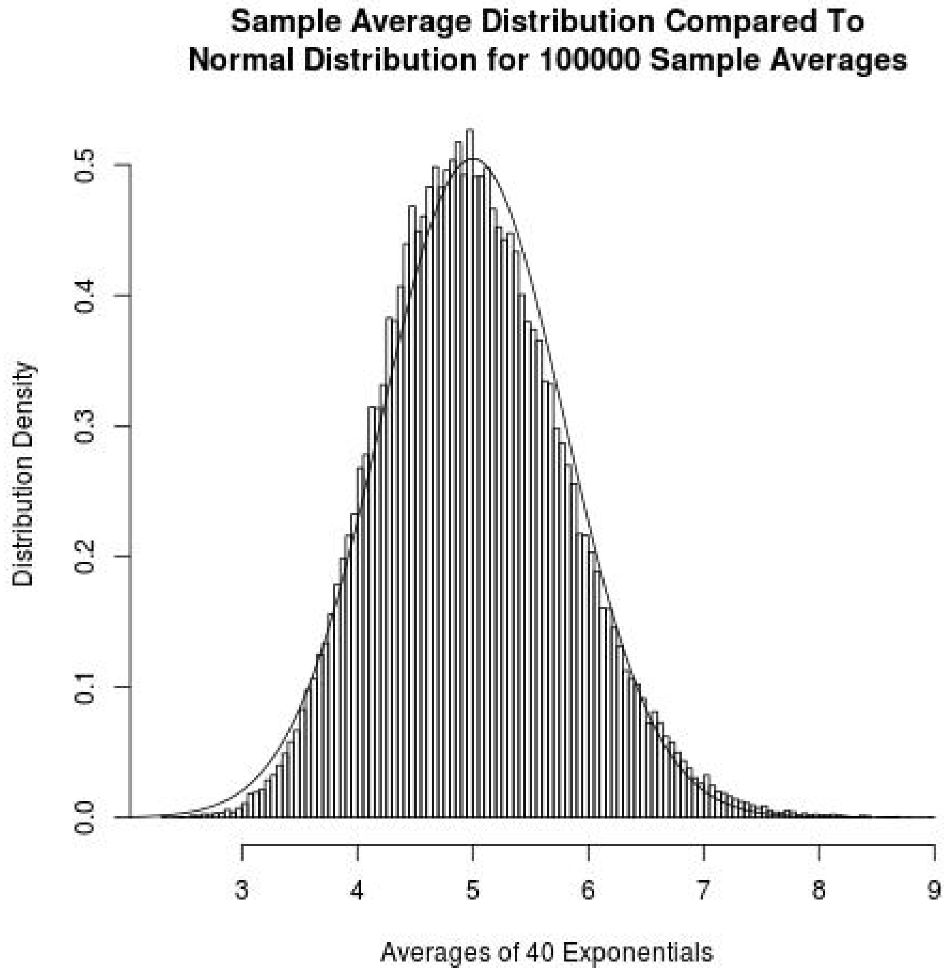
Sizes	1000 Samples	10000 Samples	100000 Samples
Mean	5.02099757521223	5.0042869577768	5.00277496190819
Variance	0.660136161503146	0.638683943765029	0.627799755419409

These values are taken from the R code output; the entirety of the statistics report can be seen at the bottom

In these histograms you can see the distribution of the sample averages more closely approximating a normal distribution as the number of sample averages increases. This is expected as per the Central Limit Theorem and Law of Large Numbers




Sample Average Distribution Compared To
Normal Distribution for 10000 Sample Averages
Distribution Density
0.0
0.1
0.2
0.3
0.4
0.5
3
4
5
6
7
8
9
Averages of 40 Exponentials


Sample Average Distribution Compared To
Normal Distribution for 100000 Sample Averages
Distribution Density
0.0
0.1
0.2
0.3
0.4
0.5
3
4
5
6
7
8
9
Averages of 40 Exponentials

```
1000 samples of the average of 40 exponentials with lambda=0.2 were taken:

Population Average: 5
Theoretical Mean of Averages of 40 exponentials: 5
Experimental Mean of Averages of 40 exponentials: 5.02099757521223

Theoretical Variance of Averages of 40 exponentials: 0.625
Experimental Sample Variance of Averages of 40 exponentials: 0.660136161503146

Theoretical Standard Devation of Averages of 40 exponentials: 0.790569415042095
Experimental Standard Devation of Averages of 40 exponentials: 0.812487637754044

10000 samples of the average of 40 exponentials with lambda=0.2 were taken:

Population Average: 5
Theoretical Mean of Averages of 40 exponentials: 5
Experimental Mean of Averages of 40 exponentials: 5.0042869577768

Theoretical Variance of Averages of 40 exponentials: 0.625
Experimental Sample Variance of Averages of 40 exponentials: 0.638683943765029

Theoretical Standard Devation of Averages of 40 exponentials: 0.790569415042095
Experimental Standard Devation of Averages of 40 exponentials: 0.799177041565277

100000 samples of the average of 40 exponentials with lambda=0.2 were taken:

Population Average: 5
Theoretical Mean of Averages of 40 exponentials: 5
Experimental Mean of Averages of 40 exponentials: 5.00277496190819

Theoretical Variance of Averages of 40 exponentials: 0.625
Experimental Sample Variance of Averages of 40 exponentials: 0.627799755419409

Theoretical Standard Devation of Averages of 40 exponentials: 0.790569415042095
Experimental Standard Devation of Averages of 40 exponentials: 0.792338157240587
```